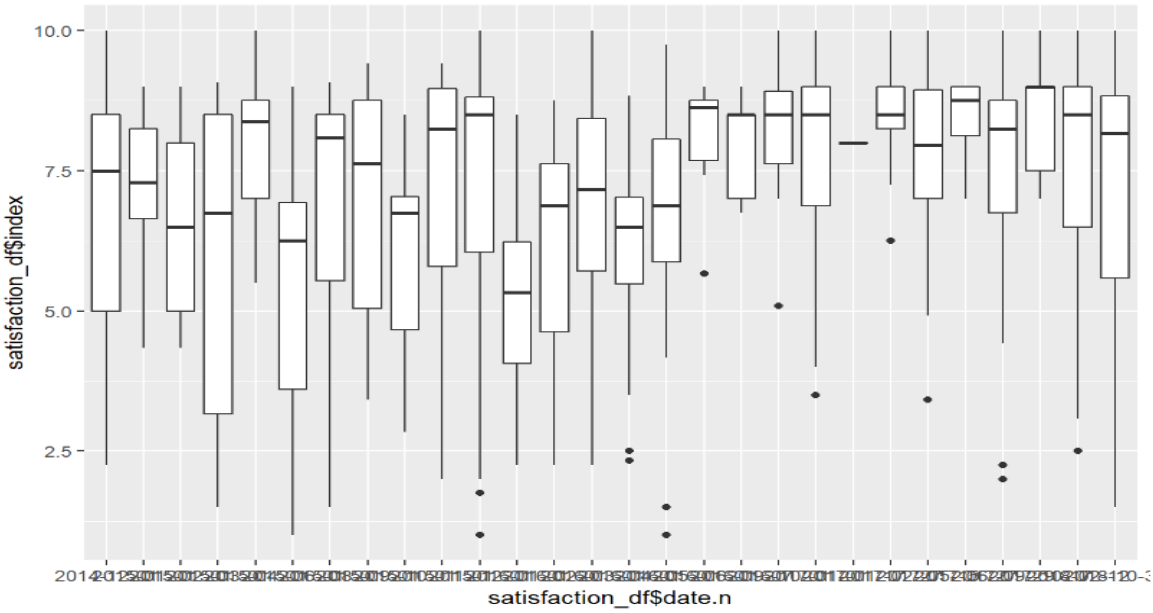
Annex 1 – Analysis and results

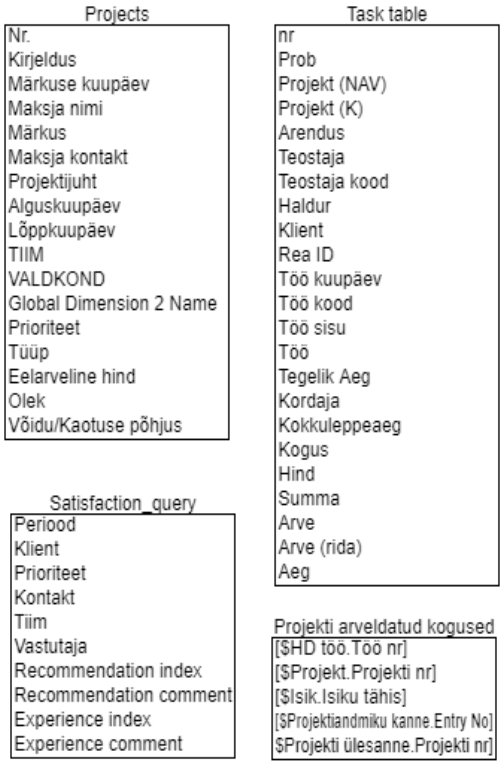
#1 Overall satisfaction over period of time

RECOM	exp	**index**
Min. : 0.000	Min. : 0.000	Min. : 1.000
1st Qu.: 7.000	1st Qu.: 7.000	1st Qu.: 6.062
Median : 8.000	Median : 8.000	Median : 7.417
Mean : 7.502	Mean : 7.514	Mean : 7.057
3rd Qu.: 9.000	3rd Qu.: 9.000	3rd Qu.: 8.604
Max. :10.000	Max. :10.000	Max. :10.000
personell	**process**	**result**
Min. :1.0	Min. :1.00	Min. :1.000
1st Qu.:3.0	1st Qu.:2.00	1st Qu.:3.000
Median :4.0	Median :3.50	Median :4.000
Mean :3.5	Mean :3.18	Mean :3.266
3rd Qu.:4.0	3rd Qu.:4.00	3rd Qu.:4.000
Max. :5.0	Max. :5.00	Max. :5.000

Overall view of responses in time:



#2 Extracted data original field description overview in sets



#3 Cleaned and modified initial dataset for project

Column name	Descriprion	Values
project code	Unique identifier of project	Random code

project type	Type of implementation project; not present in extracted dataset. Categorization is based on project description.	New implementation Maintenance /support Business requirement change Platform upgrade
Implementation methodology	method, that is chosen to implement or customize solution. Implemetation methology was added to extracted dataset based on project activities.	Waterfall V-model (or any similar combined method between waterfall and agile) Agile
price of h	sum of task and sum of requirement can be calculated if needed	
Implementation type	Not in initial dataset, implementation type was added based on solution	Full implementation ICV (ERP is accompanied by specialized modules for business line, making it simpler to implement); partial (some activities, such as CRM or POS etc) is implemented to other system and interfaced, postmodernERP (ERP is core of infosystem and best of breed software is implemented to back it up)
Start date		
End date		
Client		
Manager		

#4 Modified final task data set fields
Task data:

Column name	Descriprion	Values
task nr	Unique identifier of task	
Actual time	Time spent on task	
job nr	Unique identifier for job (in project each job represents set of activities or separate requirement)	
agreed time	Time agreed on filling the requirement	
client code	Each code corresponds to individual solution	
performer code	Each person is given separate code	
managers code-	Each client or project has single manager appointed	
Date performed	Time task was performed	DD-MM-YYYY format date
Time completed	Job completed	MM-YYYY format date
Invoiced		Y/N

Task type	Each job is recorded in helpdesk system as category. Inital dataset was reviewed based on	PM (project management); ANA (analysis; including business analysis); CON (consultation of client, change of setup etc) DEV(development); GAR (warranty – development of setup, that did not meet requirement); SISE (additional time spent on task, that is not covered by customer) TUGI (support); EDU (training of clients);
BA task type	Business analytics/Analyst tasks are divided to major groups of activities. Since there is not sufficient information about detailed activities such as time spent on elicitation with stakeholders etc it is easier to divide info into larger groups based on core activities.	Planning- **PLAN** **RES** – research and information gathering. **WS**- workshops with client. **EXP** - prototyping. **COL** - Collaboration. **LCM** -Life-cycle management **STA** -Strategic analysis – **REQ -** Requirement - **DIS -** Design - **SEV -**Solution evaluation, testing

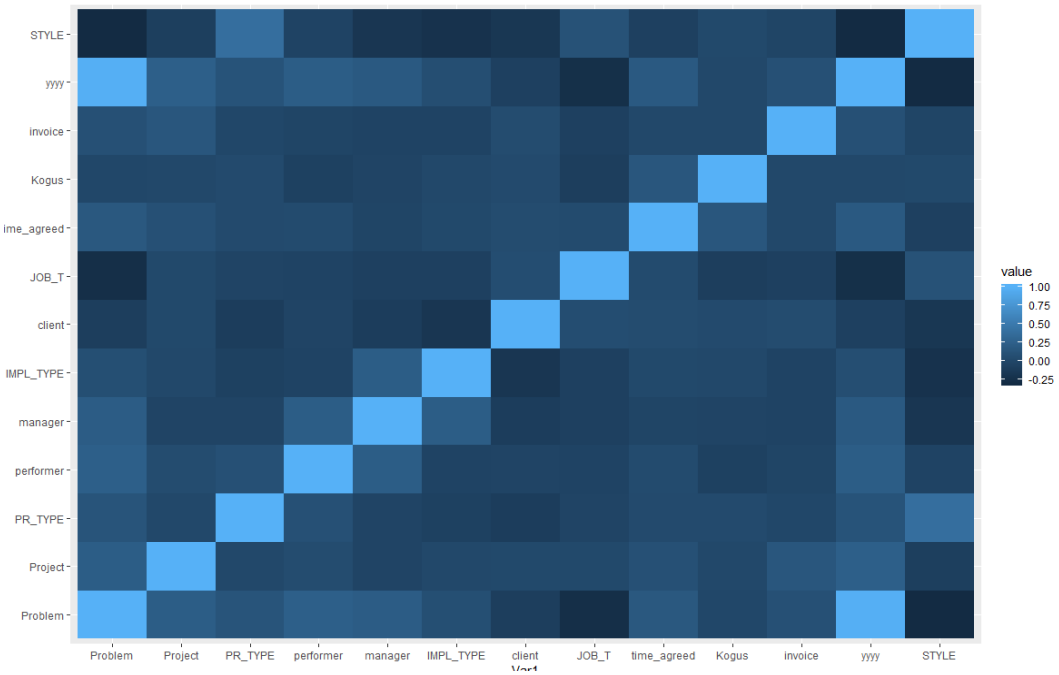
#5 Data set correlations

In order to spot any possible correlations in data-set, finalized data-set was tested to see any significant correlations between value sets. In order to create correlation matrix in R it is essential, that all comparable data is in number format. To test correlations, standard cor() function was used. Function itself creates matrix that calculates correlations between variables in observations.

In first order the correlations were used to get initial information if activites have any significant relation.

As expected, there are correlations between project and client and also between problem and year.

Rest of complete data-set has no significant correlations.



In order to analyse it further several subsets were created based on correlating data, such as job_type and implemention type, as differences in processes suggested.

	Var1	Var2	value
1	Problem	Problem	1.00
2	Project	Problem	0.22
3	PR_TYPE	Problem	0.13
4	performer	Problem	0.24
5	manager	Problem	0.21
6	IMPL_TYPE	Problem	0.08

To do further analysis, all logically correlated data was removed. Since client is directly linked to project and through that also account manager and task, it is useful to eliminate client from further analysis data-sets There is also direct correlation between ordinal numbers (such as project code and and date), therefore date was also removed from further analysis.

In order to view possible trends, separate data frames were created for each year works were performed and each type of tasks, that is possible. Each data-set with work type was tested using same logic.

By correlating initial non ordinal data it was discovered, that there are no significant correlations between types of work and style of implementation. As expected, there are some correlations between project based variables, such as project type and implementation style, what is clear indicator, that data-set is complete .

#6 Measures

all tables are displayed as **five-number summary** from R function Summary

Measure	Ratio of warranty works performed in projects by quantile of Business Analysis					
Description	.....					
Quantile of BA	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Q1	0.07	1.22	3.4	9.28	10.98	93
Q2	0.5	2	6	23.39	17.5	403.5
Q3	0.25	1.75	5.5	16.46	15	97.75
Q4	0.5	1	2	2.89	4.19	8
No BA	0.25	0.75	1.5	4.07	3.5	20.5
Observation	Each next quantile is decreacing in all values, however it seems, that projects with no BA activities have slightly lower values.					
Importance	ana_gar_r					

During proccess analysis data was divided to 4 different category based on listed evaluation using quartile distribution and business analysis activities are divided to datasets.

After initial dividing resulted dataset was applied additionally to Projects to compare ratios and works, that were labelled warranty.

Measure	Ratio of warranty works performed in projects by quantile of Business Analysis					
Description	Each project has one or more issues, that has to be redone (misunderstood requirement, insufficient testing, mistakes in code ect) during or after go-live. Projects with no warranty works are disgarded.					
Presence of BA	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Q1_warranty_h	0.5	1	3.75	20.79	13	302
Q2_warranty_h	0.5	2	6	23.39	17.5	403.5
Q3_warranty_h	0.25	1.75	5.5	16.46	15	97.75
Q4_warranty_h	0.5	1	2	2.89	4.19	8
No BA warranty	0.25	0.75	1.5	4.07	3.5	20.5
Observation	Each next quantile is decreacing in all values, however it seems, that projects with no BA activities have slightly lower values. Statistically it is relevant, however					
Importance						

Measure ana_gar_r	Ratio of analysis work compared to warranty works performed in projects by quantile of Business Analysis .					
Description	Measure is obtained by dividing all hours of Business Analysis to total hr of warranty works					
Quantile of BA	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.

Q1	0.07	1.22	3.4	9.28	10.98	93
Q2	0.44	3.92	9.63	18.78	19.32	260
Q3	1.13	5.84	20.61	33.95	34.74	274.33
Q4	4.12	9.7	40.04	126.22	151.75	951.25
Observation						
Importance						

To get results on customer satisfaction following results were revieved.

Measure sat_3Q$index	Customer satisfaction responses (440 observations on projects) were divided to quantiles based on index of answers:					
Description	Index of given responses on set of quantiles thereby dividing set to four					
Quantile of score	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Q1	1	2.77	4.42	4.01	5.17	6
Q2	6.08	6.5	6.75	6.77	7	7.33
Q3	7.42	8	8.38	8.25	8.5	8.58
Q4	8.67	8.83	9	9.17	9.42	10
Observation						
Importance						

To get better overview additional business anaysis ratios were created inside each quantile projects

Measure ratio_q1$ana_tot_r	Hours of analysis performed divided by total hr of project						
Description	Ratio of analysis to total work compared in satisfaction quantiles. Normalized by 100						
Quantile of score	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Q1	1,12	11,64	16,4	24,19	24,49	100	7
Q2	0,61	11,9	16,31	25,83	29,19	100	7
Q3	2,24	9,26	14,23	24,05	25,18	98,8	9
Q4	1,96	9,4	16,1	22,84	26	100	8
Observation							
Importance							

Measure ratio_q1$ ana_dev_r	Hours of analysis performed divided by time spent on developments of project						
Description	Ratio of analysis to development compared in satisfaction quantiles.						
Quantile of score	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Q1	0.09	0.39	0.82	3.96	1.27	125	21
Q2	0.05	0.34	0.81	3.15	1.26	118.63	20
Q3	0.05	0.26	0.66	2.78	1.33	68.5	23
Q4	0.07	0.34	0.59	1.42	1.24	22.13	21
Observation	Combined with last measure this shows correlated view on business analysis and development. The two tables combined show, that customer satisfaction is not related to amount of business analysis,.						

Measure ratio_q1$ANA	Total hours of business analysis performed by project						
Description	Total amount of analysis in satisfaction quantiles.						
Quantile of score	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Q1	0.5	7	35.25	100.74	125	951.25	7
Q2	0.5	6.5	31.25	99.59	131.88	1711	7
Q3	0.5	9.13	40.75	127.46	126.75	1711	9
Q4	0.5	12.5	46	131.3	127.6	1711	8
Observation	Combined with last two measures it is possible to conclude, that based on nature of survey data collection, customer satisfaction is not related to business analysis performed.						
Importance							

To test if there are changes in different ratios, the BA activity set was applied to project task and projects were divided to quartiles and tested against quantiles of satisfaction.

Measure Q1$sat	Ratio of all work compared to warranty works performed in projects by quantile of Business Analysis .						
Description	Measure is creatd by calculating score of satisfaction ratio from each quantile of business analysis set and using same logic on non-ba set.						
Quantile of BA	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Q1	2.33	6.38	7.58	7.23	8.5	10	99
Q2	1	5.4	7	6.43	8.5	9.08	80
Q3	2.25	5.06	6.58	6.51	8.25	10	83
Q4	1	4.96	7.08	6.71	8.75	10	100
No BA	2.25	6.38	8	7.12	8.5	10	361
Observation	There are some inconsistent patterns, that allow to conclude, that with this survey method it is not possible to link ratings to types or amounts of work.						
Importance							

7 Time comparison

To compare agreed time with actual time spent dataset was modified to summarize all time spent on task with relation to project. Since time agreed is marked on issue (requirement) level, in first order the quantiles were applied to task level dataset.

After quantile description was added, all observations actual time were summarized to issue level, since agreed time is set on issue level. After that all duplicate issue numbers were eliminated from dataset and not nessesary fields were removed.

Then all data was aggregated to project level by summarizing all work time actually performed and all time agreed and additional ratio was added. Choosen ratio was actual time spent compared to time agreed. Since number of issues are without agreed time, this ratio allows to get best insight on project level overall performance.

Results of different sets were as:

Quantile	Min	1st Q	median	mean	3rd	max	comments
1Q ANA	0.42	1.31	2.2	13.99	10.42	336	
2Q ANA	0.45	1.95	4.06	15.1	10.42	482.75	
3Q ANA	0.56	1.69	3.46	11.65	8.01	510.25	
4Q ANA	0.23	1.26	2.47	6.35	5.01	85.5	
No ANA	0	0.92	1.25	2.92	2.21	84.39	Set consists mainly on support and training projects

#8 By comparing project dataset's project end date with task dataset's latest marked date following results were received:

> Duration difference in 1-st quartile of business analysis set

Median Mean

62.0 148.1

> Duration difference in 2-nd quartile of business analysis set

Median Mean

130.0 106.4

> Duration difference in 3-rd quartile of business analysis set

Median Mean

-110.50 235.99

> Duration difference in 4-th quartile of business analysis set

Median Mean

78.0 120.2

> Duration difference in set with no business analysis

Median Mean

- 52.0 158.1

This allows to conclude, that presence of BA has effect on final timeline, however the differences are not significant and are most likely based on projects, rather than.

#9 Warranty works:

After initial dividing resulted dataset was applied additionally to Projects to compare ratios and works that were labeled warranty.

Presence of BA	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Q1_warranty_h	0.5	1	3.75	20.79	13	302
Q2_warranty_h	0.5	2	6	23.39	17.5	403.5
Q3_warranty_h	0.25	1.75	5.5	16.46	15	97.75
Q4_warranty_h	0.5	1	2	2.89	4.19	8
No BA warranty	0.25	0.75	1.5	4.07	3.5	20.5

P-values were calculated as follows:

data: Q1 and Q2
df = 111, p-value = 0.08265
data: Q1$and Q3
df = 101, p-value = 0.002728
data: Q1 and Q4
df = 64, p-value = 0.0006028
data: Q2 nd Q3
df = 118, p-value = 0.06721
data: Q2 and Q4
df = 81, p-value = 0.0003593
Q3 and Q4
df = 71, p-value = 0.006116

#10 Ratios of works- business analysis time ratio to warranty works time

> analysis to warranty ratio
Median Mean
3.40294 9.28456
> summary(Q2$ana_gar_r)
Median Mean
9.6250 18.7792
> summary(Q3$ana_gar_r)
Median Mean
20.614 33.951
> summary(Q4$ana_gar_r)
Median Mean
40.037 126.219

> Warranty to total ratio Q1 business analysis batch
Median Mean
0.01613 0.03653
> Warranty to total ratio Q2 business analysis batch
Median Mean
0.01207 0.03252
> Warranty to total ratio Q3 business analysis batch
Median Mean
0.00957 0.02498
> Warranty to total ratio Q4 business analysis batch
Median Mean

0.00913 0.02012

> Warranty to total ratio no business analysis batch
Median Mean
0.0476 0.0569

#11 Ratios of works within different implementation methods:

> Analysis to total ratio Waterfall
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.00442 0.10063 0.17788 0.26060 0.34101 1.00000 38
> Analysis to total ratio Agile
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.00619 0.08009 0.17949 0.26893 0.35635 1.00000 90
> Analysis to total ratio V-Model
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.00241 0.09079 0.14672 0.19747 0.22649 1.00000 62

> Development to total ratio Waterfall
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.00473 0.13151 0.24301 0.27826 0.38243 1.00000 109
> Development to total ratio Agile
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.0035 0.2330 0.4313 0.4173 0.5714 1.0000 81
> sDevelopment to total ratio V-model
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.00681 0.12713 0.23032 0.25960 0.36668 1.00000 97

12 Time comparison

To compare agreed time with actual time spent dataset was modified to summarize all time spent on task with relation to project. Since time agreed is marked on issue (requirement) level, in first order the quantiles were applied to task level dataset.

After quantile description was added, all observations actual time were summarized to issue level, since agreed time is set on issue level. After that all duplicate issue numbers were eliminated from dataset and not nessesary fields were removed.

Data was aggregated to project level by summarizing all work time actually performed and all time agreed and additional ratio was added. Chosen ratio was actual time spent compared to time agreed. Since number of issues are without agreed time, this ratio allows to get best insight on project level overall performance.

Results of different sets were as:

Quantile	Min	1st Q	median	mean	3rd	max	comments
1Q ANA	0.42	1.31	2.2	13.99	10.42	336	
2Q ANA	0.45	1.95	4.06	15.1	10.42	482.75	
3Q ANA	0.56	1.69	3.46	11.65	8.01	510.25	
4Q ANA	0.23	1.26	2.47	6.35	5.01	85.5	
No ANA	0	0.92	1.25	2.92	2.21	84.39	Set consists mainly on sales, support and training projects

By comparing project dataset's project end date with task dataset's latest marked date following results were recieved:

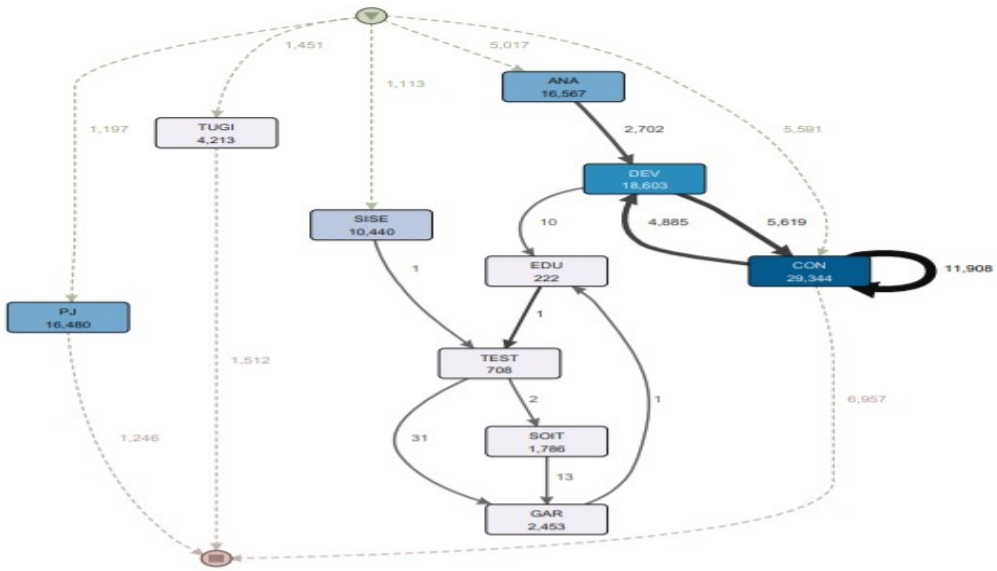
```
> summary (as.integer(Q1_date$differ_duration))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
-1168.0   -19.0    62.0   148.1   327.0  1810.0       2
> summary (as.integer(Q2_date$differ_duration))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
-1764.0   -38.5   130.0   106.4   391.5  1153.0       2
> summary (as.integer(Q3_date$differ_duration))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
-500.00   -4.75  110.50  235.99  401.50 1649.00       2
> summary (as.integer(Q4_date$differ_duration))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -757.0   -10.0    78.0   120.2   248.0  1050.0
> summary (as.integer(NBA_date$differ_duration))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -642.0    12.0    52.0   158.1   185.0  1910.0
```

this allows to conclude, that stronger presence of BA has no significant effect on final timeline.

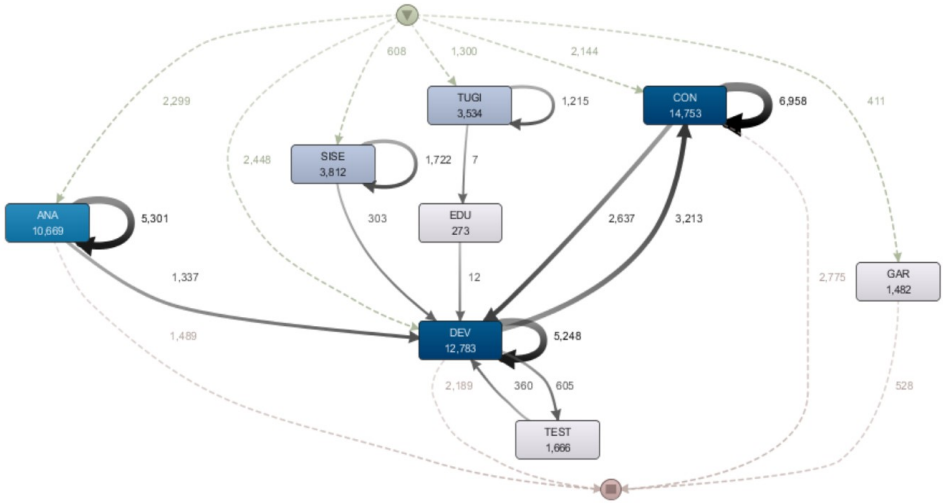
#13 Process mining has purpose of finding actual processes based on logs. In thesis differences between methods were used. Since all different types of projects, implementation methodologies and implementation types were extracted to separate datasets using R and exported in csv format.

Here are presented project types process charts, that are not relevant to thesis.

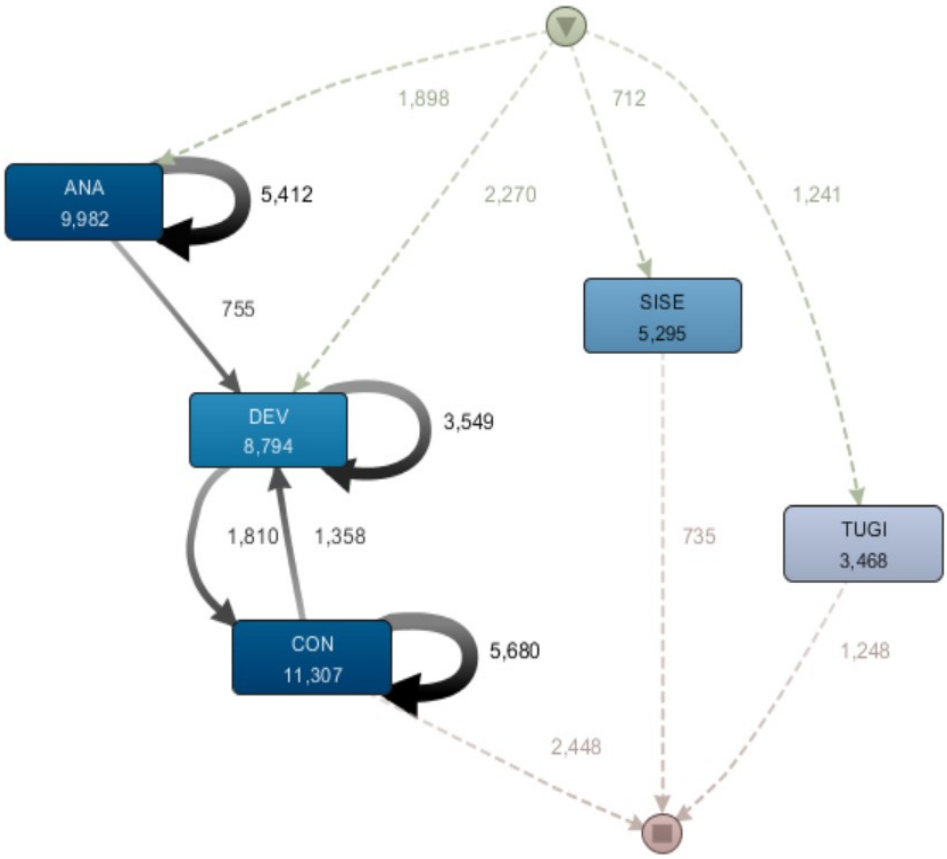
Mining result 80% on maintenance



Mining result on Business Proccess Change -> process starts with analysis that is usually separated on different task, after analysis is completed, new task is created, where testing is recorded.



Mining result on new implementations-> it is identifiable, that the process starts with analysis, then moves to development and after that to consultant. In 80% level the main process remains similar. Process was also tested in each quantile of BA activities, but it remained still the same.



#14 Implementation style based differences in satisfaction and delay

> satisfaction in Waterfall

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   5.417   7.000   6.723   8.500  10.000
```

> satisfaction in Agile

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.500   6.500   8.000   7.261   8.792  10.000
```

> satisfaction in V-Model

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   5.000   7.000   6.523   8.500  10.000
```

> Duration difference in Waterfall

```
Median    Mean
  70.0   161.2
```

> Duration difference in Agile

```
Median     Mean
```

67.00 140.10

> Duration difference in V-Model
Median Mean
245.00 281.64

> Delay in Waterfall
Median Mean
96.0 195.1

> Delay in Agile
Median Mean
94.0 221.5

> Delay in V-Model
Median Mean
56.0 171.3